

汉语词汇定量研究的运用及其特点

——兼谈《语言学方法论》的定量研究观

苏新春

(厦门大学 中文系,福建 厦门 361005)

摘 要:定量研究的兴起是当代汉语语言学一个引人注目的特点。定量分析方法在汉语词汇学中推广运用的历史背景及使用特点是:汉语研究史中“例不十,不立法”传统的继承与光大;以专书研究为计量研究崛起的主要标志;以词量为定量研究的主要内容;由手工统计向语料库使用过渡。词汇计量研究要注意选材的典型性、标注的多角度与普遍性,切入语料的角度要准,要以理论指导定量分析。定量研究方法不全等同于实验方法。

关键词:定量研究方法;汉语词汇学;数据库;统计分析

中图分类号:H03 **文献标识码:**A **文章编号:**0438-0460(2001)04-0135-08

一、汉语词汇计量研究的形成与运用

语言研究中的定量方法就是通过对语料进行数的反映,以达到认识语言规律和特点的做法。80年代以来,加强定量研究的倡导在汉语词汇学界得到很好的呼应。第一部有着严格统计学意义的反映词量、词长、词汇分布、词语构成等断代词汇状况的数据词典《现代汉语频率词典》于1986年正式出版。[1]先后还出版了一些大型词汇统计数据资料书。[2][3]但定量研究方法对汉语词汇研究产生更大影响的不是几部词汇数据词典的出版,而是它在专业工作者的实践工作中日益受到重视,并得到应用。几千年来汉语词汇研究传统,都是以具体词语词义为主要对象,以考释为主要目的,以研究者的主观感悟为主要手段。到现代,虽然重视了对词汇整体的理论属性的探讨,逐渐摆脱了专注于具体词义的考释性研究的旧格局,但在研究手段上却一直没有大的改变,靠的仍是研究者个人的语感,个人所熟悉的部分语料。因此,定量研究方法的引进与推广,在当代词汇研究中有重要的革新意义。

汉语词汇定量方法的应用与发展,有着特定的学术背景与表现形式:

1. 对“例不十,不立法”传统的继承与光大

汉语言学界一直流行着“让材料说话”的观点,特别强调对语言材料的充分占有。不说空

收稿日期:2001-06-07

作者简介:苏新春(1953-),男,江西南昌人,厦门大学中文系教授。

话,不作无根底的文章,成为长期以来评判语言学论文的一条不成文的,却几乎至高无上的标准。“让材料说话”铸成了汉语言学界特有的求实之风。在20世纪中前期,它一直是对汉语研究者最具影响力的因素之一。在这种思想的指导下,在语料挖掘与理论概括之间有着一句流传面很广的名言:“例不十,不立法”。王力先生对此很推崇,把它作为汉语史研究的基本原则之一加以论述:“所谓区别一般和特殊,那是辩证法的原理之一。这里我们指的是黎锦熙先生所谓的‘例不十,不立法’。我们还要补充一句,就是‘例外不十,法不破’”。[4](P19)这里的“十”不是定量,只是从谨慎立论的角度来说,却含有明显的以“量”取胜的意思。不以孤证立论,根据语料的多少立论,这就是汉语研究中最朴素的量的观念,它已成为中国语言学研究中的一种潜意识与自觉行为。

2. 以专书研究为计量研究崛起的主要标志。

言语材料的无限,使得语言研究要实行量化总会面临难于舍取的尴尬,这就很自然使人们转而考虑选取既容纳有相当言语材料,又有一定时代和语言特征的专书。符合这两个要求的当然首先是经过历史检验的汉语史各个阶段的书籍。因此,重视专书研究的做法首先在汉语史学界流传开来也就是很自然的事了。《左传》、《史记》研究专家何乐士先生的一段回忆,颇为真实地记载了这一风气之端起:“数十年我们的师辈一直不间断地倡导专书研究。拿我亲身感受而言,60年代初,陆志韦先生就亲自带领我们投身这项工作。吕叔湘先生在担任语言所所长期间曾不止一次对古汉语研究室的研究人员强调,汉语史研究应以专书研究作为基础。1979年他在写给古汉语研究人员的一份建议书中写道:‘要对古代汉语进行科学的研究,就要注意时代和地区的差别。对这些差别,现在还只有一些零碎的认识,还缺少系统地探索的成果。要进行研究,现在还只能先拿一部一部的书做单位,一方面在同一作品中找规律,一方面在作品与作品之间就一个个问题进行比较。’丁声树先生生前也一再教导我们说,那种任意选取例句的做法不能科学地总结规律,应对专书进行穷尽的调查研究。”[5]

专书研究的风气真正刮起来是在20世纪的80年代。这时各个历史阶段的专书都陆续有人进行了专门的研究,他们努力于反映该书的词汇整体面貌,如何乐士的《左传》、《史记》研究;[6]张双棣的《吕氏春秋》研究[7];毛远明的《左传》研究[8]等。张双棣先生描绘了《吕氏春秋》的词汇概况:“全书有单音词2972个,复音词2017个,总共近五千词。单音词中,名词最多,达1371个,动词次之,有1298个,形容词又次之,有464个,其他类词共有272个。”[7]这已经是相当严格的定量研究了。尽管学者们大多依靠的是手工操作,语料统计数字还难保绝对的准确,但从数量的多少来看词汇的结构规律,已成为一条基本准则。

近代汉语、敦煌学、佛教词汇学、语用学、社会语言学等相关学科中展开的其它词汇专题的定量研究,成果蔚为可观。这些专题性的词汇定量研究论文都采用了统计的方法。如程湘清先生的《先秦双音词研究》:“上述词语共615个,其中最多的是指人,指事物、指时地的名词语,共386个,占全部词语的62.76%;其次是指动作、行为或变化的动词语,共139个,占22.6%;再次是指人和事物的性质、状态的形容词语,共90个,占14.64%。从结构上看,绝大多数是运用语法手段词序的特点构成的,共586个,其中居首位的是并列式,共307个,占52.4%。”[9]程文作于80年代初,他所作的定量工作在当时算是相当彻底的。

定量研究的方法,对词汇研究有其特别的便利之处。因为词汇的独立性比较强,形式化工作做起来比较容易,定量的结果有形可感,容易为人们所接受。因此,不断有学者特别呼吁加

强词汇的定量研究。“如果不作定量分析,就很难把握住汉语诸要素在各历史时期的性质及其数量界限。我们的断代描写和历时研究也必然要陷在朦胧模糊的印象之中。从随意引证到定量分析,是古汉语研究为走向科学化而迈出的重要一步”。[10]

3. 以词量为定量研究的主要内容。

现有的词汇定量研究,涉及到的内容已相当广泛。这里的定量研究主要表现为词量的研究。词量指的就是词语在结构、单位、分布、使用等静态和动态中表现出来的数量、频率、范围的多少、高低、广狭,如词语数的量、构词能力的量、使用高低的量、结构方式多与少的量、语境语体的分布量等。词量的研究在词汇研究中具有重要作用。语言的种种重要性质和特点,都会通过“量”上的高频、高见反映出来。愈是重要的、常见的词汇现象,愈是会有较多的“量”来体现它。

4. 由手工统计向语料库使用过渡

当前以词量为主要内容的定量研究,大多数学者主要是靠手工摘记卡片的方式。这与研究者的知识背景及掌握工具与手段的能力有密切关系。随着时代的演化,研究者知识结构的改变,特别是电脑的普及,现在依靠语料库和统计软件来完成的词汇研究已经逐渐出现。如《九十年代汉语词汇地域分布的定量研究》就在六百万字的汉语词语库的基础上,通过流通量与分布率的定量分析,论述了中国大陆、台湾、香港三地词汇使用的同异程度。[11]目前汉语词汇研究中运用统计学定量研究,开展得较好的是在方言词汇研究领域和语料库领域。

在专题语料库领域,据朱小健先生的统计,在中国大陆建成的专书或封闭的专题语料库已有近10个。[12]笔者近两年来建设的“《现代汉语词典》数据库”,把这部有相当权威性、规范性的语文词典作为现代汉语词汇总貌的一个缩影进行计量分析,已完成了10余个专题的研究,对词汇学和词典学上的一系列问题进行了定量的分析研究。建立封闭性的专题词汇语料库,小至某个专题、某本专书,大至某个断代,对于深化词汇研究,提高词汇研究的准确性和系统性,都是极有意义的。这项工作已开始吸引了愈来愈多的词汇研究者投身其中,最近已有人在着手建设近代汉语的断代词汇语料库。

在方言学领域,已有不少论文运用统计学上的相关系数方法分析方言词汇或方言语音,以达了解方言分区之间亲疏关系的目的。尽管它们所用的方法不尽相同,有的还停留在对方方法本身的摸索,改进层面,但相比于其他领域的词汇定量研究,已经领先了一大步,显示出了统计学方法在语言研究中的广泛应用前景。

语料库领域的词汇计量研究,除了来自汉语词汇学界之外,还有一股力量值得特别注意,就是来自中文信息处理界的研究者。他们在进行汉语的自然语言处理时,一般都建有千万、亿

正式刊出的有:《当代汉语变化与词义历时属性的释义原则——现汉二版、三版“旧词语”对比研究》,《中国语文》2001年第2期。《同形词与“词”的意义范围——析现汉的同形词词目》,《辞书研究》2001年第5期。《现代汉字的范围及其属性标注》,《汉字文化》2001年第1期。《“人”“机”分词差异及规范词典的收词依据——对645条常用词未见于现汉的思考》,《辞书研究》2001年第2期。《比喻义的训释与比喻义的形成》,《杭州师范学院学报》2001年第5期。《现代汉语词典对异形词的整理及对当前规范工作的启示》,《语言文字应用》2001年第3期。

徐时仪:《数字化建设与断代词典编纂》,第二届全国中青年辞书学工作者学术研讨会会议论文,2001年4月24日,上海。

万字规模的大型语料库。前者重在对封闭语料计量后的精加工,多深入到词义内在规律的探讨,后者则重在对海量词汇的词形研究,他们对数据库的高效率利用,往往能对词汇形式与结构各个方面的量作出很周全精确的分析。这股力量还未引起汉语词汇界的足够重视,相信一旦与汉语词汇研究界结合起来,将大大有助于汉语词汇理论研究的深入发展。

定量方法在汉语词汇研究中已显示出极强的生命力,必将对由来已久的定性式研究产生巨大的震动和推进,彻底改变以往那种凭作者语感,靠对语料摘取式的主观色彩浓厚的做法。它具有自己的鲜明特点:肯定受到西方语言学定量方法的影响,但更多的是对自身研究传统中数量观念的继承与提升;词汇定量主要靠的是算术统计法,反映的多属数值与比率的直观关系,使用复杂的数学计算还不多,推理、间接的数值关系探讨也很少;针对汉语史中词汇问题研究的多,现代词汇研究的少;面向专书词汇的多,面向断代词汇整体面貌的少,面向辞典词汇的更少;选取容易封闭的书面语多,选取开放性的口语少。这就是汉语词汇定量研究的现状。

二、汉语词汇计量研究中要注意的若干问题

围绕词量这一核心问题展开的词汇定量研究,有以下几点需要加以注意。

1. 选材要有代表性、典型性、封闭性

定量研究对语料的选取有着很高的要求,不同的语料在定量研究中会显示出不同的价值。如在方言近似率的定量研究中,是选取最高频的常用词,还是选取数量在几倍之多的一般词语;是以词为单位,还是以语素为单位,就有着完全不同的意见。[13]在大型的词汇统计中,语料选取是否科学往往决定成败。这也就是《汉字频率表》之所以能在同类字表中后来居上,可信度高的一个重要原因。[14]在对现代汉语词汇进行总貌性的研究中,我们选取了《现代汉语词典》作为建立数据库的材料,首要考虑的因素也就是它的语文性、规范性、普遍性与权威性。语料具有了代表性、典型性,才使科学结论的提出有了可靠的前提。

定量研究对语料的另一个要求,是语料的封闭性。只有封闭才能做到定量统计的穷尽、准确。建立《现代汉语词典》数据库时,发现同一个版本在不同印刷次数时,常出现了局部的改动,或改释义,或增删词目。尽管这些挖版式的改动不涉筋骨,但对定量研究来说,毕竟会时不时带来拂不去的遗憾。这使我们在建立数据库的过程中,曾一度考虑光引用“版本”的概念可能还不行,还得加上“第几次印刷”。当因种种原因语料难以做到封闭时,随机采取就不失为一个好办法。笔者几年前曾手工做过《现代汉语词典》的单、复音词载义量的对比。在全书中抽取了相隔固定页码距离的16页共480条复合词作调查,发现每个词的平均义项数是1.308个。[15]这个结果离“大多数词语是多义词”的传统说法离得太远。现在利用数据库来统计,发现全书45606条复音词的平均载义量是1.163,与原研究结论相去不远。可知随机取样得当,同样能收到全部语料封闭调查的效果。

李兆麟先生认为:“《汉字频度表》由于取样不恰当,选了‘文革’时期报刊上的许多文章作统计材料,科学技术书刊总字数在全部统计材料中所占的比重较大……”。“《汉字频率表》在统计时所用的语言材料及各类语言材料所占的比例是最恰当的,具有代表性、多样性和均匀性。”详见“参考文献”[17]。

如1996年的《现代汉语词典》(修订本)在1998年以后的印刷中在901页就新增了“母老虎”一词。

2. 词汇标注的多角度与周遍性

要对词汇的语义、语音、语法关系进行多角度、多层面的定量研究,少不了要进行词汇属性的标注。这是进行深入定量研究非常重要的一环。《现代汉语语法信息处理词典》对动词属性信息的标注就达 128 项。[16]这是一部为计算机信息处理用的后台词语语料库,其实它本身又是人们对语言精心研究的结果。笔者在进行“同形词”的研究中,为了全面反映《现代汉语词典》中 1 302 个、640 组同形词的语法、语音、词义之间的同异关系,进行的标注达 20 多种。[17]所谓周遍性就是同一种标注要覆盖所有的语料,无论是有或无,或有的不同级别,都要加以标示,这样才能使语料库处理起来更为便捷、准确。

3. 切入语料的角度要准,理论融入定量分析的全过程

语料的计量分析是定量研究中非常重要的一环,甚至成为这一研究最抢眼的地方。但这只是它的外部表现,不能以为只要把语料作了量的分析就完成了所有的研究工作。语料的量化工作并不意味着自动获得研究的意义,语料不会自动显示内在的价值。它只是理论研究的一种工具。定量研究中的语料选取、语料标注、量化分析,每一个环节都要依靠理论的指导,才会使定量研究获得生命力。要克服那种只有定量,没有理论,为定量而定量,满足于统计分析过程的纯形式化研究。如《现代汉语词典》(1983 年版)有 2 482 条词含有比喻义,使用了六种不同的训释方式,这些训释方式是否使用得当,相互之间有何联系,固然很值得词典学、释义学来研究,但更有意义的是它们反映的是些什么不同性质的词义,由此把探讨的视角确定为释义方式与词义成分之间有着怎样的对应关系。从而发现《现代汉语词典》作为一部规范词典,它所反映的词义还包括了一部分富于生命力,处于形成过程,尚未定型完全的新词义成分。

以上几点有的体现了定量分析中的技术要求,而更多的是对定量分析法本身的认识。定量分析能够很好地解决语料的处理、分析,直接融入论证过程之中,直接有助于研究结论的得出,但这毕竟属于语言研究技术层面的东西。它不能代替人们对语言规律的本质认识。它只有在正确的语言理性认识的指导下才会焕发生命力。

三、实验方法不全等同于定量研究方法

桂诗春、宁春岩的《语言学方法论》出版后,一时洛阳纸贵。这当然首先是因为它内容丰富与系统,它观点的极致也格外引人注目,即特别强调了定量研究方法的运用。“54 %的人使用的是简单的思辨性的方法,随机性很大。这些研究的成果难登大雅之堂”。“难登大雅之堂”,[18](P)成为近 20 年来国内语言学界倡导计量研究之风中对非定量研究最严厉的批评。

尽管这种批评加上了一些范围限定,指的是“我国的外语工作者”;对汉语学界则表示了相当的宽容,“对母语的研究,也可依赖我们对母语的直觉观察来进行思考,提出理论模型。但是对非母语的描述性研究和实验性研究却必须以数据为依归。而我们有 80 %的研究都是不依赖数据的”。[18](P)但这种严厉批评下的宽容仍使汉语学界陷入沉思:怎样的研究才算定量研究?汉语学界有没有定量研究?不同的定量研究方法有何不同的特点与效用?

之所以会提出这些疑问和思考,是因为《语言学方法论》对定量方法有着非常严格的解释。它在“9.2. 定性方法和定量方法”作出了这样的论述:[18](P212)

定性方法

1. 自然观察
2. 现象学观点:“站在活动者本人的角度去了解人类行为。”
3. 归纳
4. 综合
5. 描述性

定量方法

- 操纵和控制
- 逻辑实证主义观点:“对社会现象事实和原因的了解无需考虑个人的主观状态。”
- 演绎
- 分析
- 推断性

书中对表中的5个区别点都一一作了详细的论述。下面试看对第1点“自然观察——操纵和控制”的说明:

自然观察

1. 观察面广,但分散
2. 变量不加控制,有利于了解它们的复杂关系,但容易顾此失彼
3. 注意内容,但容易忽略形式
4. 解释力强,但容易主观
5. 接近现实,但时间长

操纵和控制

- 观察面窄,但集中
- 变量有所控制,有利于了解它们的因果关系,但容易简单化
- 注意形式,但容易忽略内容
- 客观性强,但解释力弱
- 时间短,但人为的成分大

从这些论述可以看出,《语言学方法论》所谈的定量方法其实就是“实验方法”或“实证方法”。因为是实验方法,所以要控制各种不同的实验因素和实验条件;因为是实验方法,所以事先要有实验方案,之后要有实验报告,并进行相互印证式的检验;因为有方案在先,验证报告在后,所以它属于演绎,属于分析,属于推断。

实验方法当然属于定量方法的范畴。它考虑了研究对象较为复杂的因素及其相互关系,运用了较为复杂的原理和手段来观察、再现这些关系。它通过模拟的方法,再现语例,并能对语例进行简化、分化,从而控制、观察其中的某个部分、某个因素的变化情况。它当属于定量方法中较为复杂、深入的层次。稍深入些的实验方法都要借助于数学,利用统计学原理,利用统计软件来处理实验数据。但能否得出只有实验的方法才属于定量方法的结论,这是很值得怀疑的。因为只要对研究对象进行过“数”的系统反映,并通过“数”来再现语料的内在规律、特点,就都应该归属于定量研究的范围。这才是定量研究的真正含义。

“什么是定量分析方法呢?所谓定量方法,就是将处于随机状态的某种语言现象给予数量统计,然后通过频率、频度、频度链等量化形式来揭示这类随机现象背后所隐藏的规律性。”^[19]“定量研究方法,将所研究的现象的有关特征实行量化,然后对取得的数据进行统计学处理,得出结论。该方法兴盛的一个主要原因,它体现了现代人们所推崇的科学精神。从根本上说,定量研究方法渗透着这样一个观念:世界上一切事物不依赖人的主观意志而存在,是可以被认识的;它们的各种特征都表现为一定的量,所以,定量的方法是认识事物的科学方法。”^[2]这里对定量分析的解释是中肯的。当然,定量研究中存在着不同的类型,不同类型的定量研究有着不同的适用范围,甚至有着难易、繁简不同的区别,这是无庸讳言的。

“按照所使用的具体方法,定量研究也进一步划为(1)使用描述统计还是推断统计;(2)使用实验方法还是非实验方法;(3)使用单变项分析还是多变项分析。”^[20]“量化研究”将材料

数据化。它的三个子类分别与研究设计、统计技术、变量个数有关。量化范式一般认为,实验性设计比非实验性设计更加严格,推断性统计比描述性统计更有概括力,多变量比单变量更全面。"[21]尽管这里把定量研究中的不同类型分出了高下之别,但都把“非实验性”“描述性”“单变量”看作是定量研究中的一部分。

不同类型的定量研究不仅有着不同的含义,不同的适用范围,也有着不同的特点与短处。一般说来,算术关系反映出来的是对象之间较为简单的数学关系,它固然简单,但直观、明了,自然、真实。而实验方法中人为的作用太强,人工影响难以除掉,这就难免会影响到数值关系的客观性。如加权是统计分析中常用的一种方法,汉语学界在进行方言之间相似度的研究中就对此有过截然不同的看法。“我们主要不加权。方言定量研究结果应该是客观性的。如果加权,势必使结果带有主观性。方言内部各种特征的地位在定量研究中应该相等。”[22]有的则主张加权:“在一种方言里,有的词汇常用,有的不常用,使用频率不同的词汇对于方言之间的词汇接近率的重要性是不同的。换句话说,词频对词汇接近率的计量统计应该是很重要的参数。所以我们将以词频为基础的词汇组频率作为权数。……单音节词中的语素负载这个词的全部语义和信息,语素重要性自然最大,权数也自然最大。”[13]这里我们不想对位居两极的说法作出评断,甚至也不想作略带褒贬之分的叙述,只想说明一点:任何一种具体的定量方式,要想绝对的完美是很难的,总会或多或少带有这样那样的局限,或来自于客观,或来自于主观,或来自于过程,或来自于方法。因此,要想其中的一种方法绝对地超乎其它,是不可能的。不能因为有了实验方法的定量,就否认非实验方法的定量,甚至进一步把定量研究完全狭义地定义等同于实验方法。

汉语学界也有人持此绝对看法:“汉语方言定量研究应该定义为:根据统计学的原理对汉语方言成分所进行的量的比较。在汉语方言学中,不是所有量的说明都是定量研究。简单的量的计算不是定量研究,只有进行量的比较才是定量研究。”[22]不过,定量研究所真正具有的含义与范围,并没有因此而出现变移、缩小。如著名语言学家许国璋先生在评价何乐士先生以“计数”为特点的定量研究时,就明确把它归之于定量研究之列:“大量的劳动和细致的分析使这篇文章成为一篇扎实的语言态研究和计量语言学研究”。[23]

近年出版的一本带有一定国际区域性的很严格的语言定量研究著作《汉语计量与计算研究》,共收论文 21 篇。我们对它的研究方法作了简要调查:

国别(地区)	研究领域	撰写文本语种
中国大陆:7	普通语言理论:3	汉语:15
中国香港:5	专题语料研究:11	英语:6
日本:3	中文信息处理:7	
中国台湾:4		
新加坡:1		
美国:1		

在用汉语撰写的 15 篇论文中,研究对象为“专题语料研究”的有 8 篇,其中使用了“词数”“句数”的计数式的定量方法的有 7 篇。7 位作者分布的地区为:中国大陆 2 篇,美国 1 篇,日本 2 篇,台湾 1 篇,新加坡 1 篇。使用

了统计学原理来作定量分析的只有 1 篇。可见,“计数”分析仍是定量研究方法中的重要一种。

定量研究方法不等同于实验统计方法;定量研究包括计数研究;“计数”方法与统计学的分析方法各有作用;“计数”方法的运用与否依语料性质、研究目的而定,这些当是不争的事实。

参考文献:

- [1]北京语言学院语言教学研究所.现代汉语频率词典[Z].北京:北京语言学院出版社,1986.
- [2]刘源.现代汉语词表[Z].北京:中国标准出版社,1984.
- [3]刘源,梁南元.现代汉语常用词频词典[Z].北京:宇航出版社,1990.
- [4]王力.汉语史稿上册[M].北京:中华书局,1980.
- [5]何乐士.专书语法研究的几点体会[J].镇江师专学报,1999,(1).
- [6]何乐士.《史记》语法特点研究[A].程湘清.两汉汉语研究[C].济南:山东教育出版社,1984.
- [7]张双棣.《吕氏春秋》词汇研究[M].济南:山东教育出版社,1989.
- [8]毛远明.《左传》词汇研究[M].重庆:西南师范大学出版社,1999.
- [9]程湘清.先秦双音词研究[A].程湘清.先秦汉语[M].济南:山东教育出版社,1982.
- [10]郭锡良.1985年的古汉语研究[J].中国语文天地,1986,(3).
- [11]陈瑞端,汤志祥.九十年代汉语词汇地域分布的定量研究[J].语言文字应用,1999,(3).
- [12]朱小健.古籍整理通用系统及其中字典的编纂[J].语言文字应用,2000,(3).
- [13]游汝杰,杨蓓.粤沪普词汇接近率研究[A].邹嘉彦,等.汉语计量与计算研究[M].香港城市大学语言信息中心,1998.
- [14]李兆麟.汉语计量研究初探——兼评现代汉语频率词典[J].辞书研究,1989,(1).
- [15]苏新春,许鸿.词语的结构类型与表义功能[A].词汇学新研究[C].北京:语文出版社,1995.
- [16]俞士汶.现代汉语语法信息词典[Z].北京:清华大学出版社,南宁:广西科学技术出版社,1998.
- [17]苏新春.同形词与“词”的意义范围——析现代汉语词典的同形词目[J].辞书研究,2000,(5).
- [18]桂诗春,宁春岩.语言学方法论[M].北京:外语教学与研究出版社,1997.
- [19]唐钰明.定量分析方法与古文字资料的词汇语法研究[J].海南师范学院学报,1991,(4).
- [20]孟悦.目前我国应用语言学研究方法的调查与分析[J].现代外语,1993,(1).
- [21]高一虹,等.中、西应用语言学研究方法发展趋势[J].外语教学与研究,1999,(2).
- [22]沈榕秋.谈汉语方言的定量研究[J].语文研究,1994,(2).
- [23]许国璋.计量的语言态对比研究[J].外语教学与研究,1987,(1).

[责任编辑:贺秀明]

Application and Characteristics of Quantitative Analysis in Chinese Lexicology

SU Xin-Chun

Abstracts Quantitative analysis is a conspicuous feature in contemporary Chinese linguistics. The historical context for the promotion of quantitative analysis in Chinese lexicology and the characteristics of its application are: it helps to develop the tradition of “no conclusion with less than ten examples” which evolved through the history of Chinese language studies; its first application is known for the starting of thorough analysis of a specific book; it is mainly used in quantifying vocabulary; and it speeds up the transition from manual calculation to using computerized corpus. This paper suggests that attention be paid to the following: multidimensionality of the chosen material and marking systems; choosing a right perspective to approach the corpus; guiding our quantitative analysis with theory; avoiding equating quantitative analysis completely with experimental methods.

Key words Quantitative analysis, Chinese lexicology, corpus, statistical analysis